Hamilton College Usability Lab
Sponsor No: FA9550-10-1-0243
Final Report

Stuart Hirshfield, Leanne Hirshfield

Hamilton College

1 Introduction

With the help of DURIP funds, we created a new Usability Research lab at Hamilton College that enables us to gather a variety of cognitive and behavioral metrics from users in order to gather real time, quantitative measures of their mental states while they work with a variety of computer interfaces. The equipment purchased with the DURIP funds is listed in Table 1. Paramount to Hamilton's approach is the use of a relatively new, non-invasive brain imaging technique called Functional Near-Infrared Spectroscopy (fNIRS) to make real-time, objective measurements of users' mental states while they work with various user interfaces. fNIRS was introduced in the early 1990's to complement, and in some cases overcome practical and functional limitations of EEG and other brain monitoring techniques. Users can wear the comfortable fNIRS device in real working conditions, allowing the measurement of their brain activity while they work with a given interface. These unbiased, real-time metrics are used to augment more traditional usability tests, providing greater insight into the effects of a given interface design on target users.

Table 1: Equipment Purchased with DURIP funds for the Hamilton College Usability Lab.

b-alert wireless EEG system	1	$20,000	http://www.b-alert.com/eeg/b-alert.html
b-alert external syncing unit	1	$5,000	http://www.b-alert.com/eeg/b-alert.html
b-alert ERP analysis software	1	$5,000	http://www.b-alert.com/eeg/b-alert.html
b-alert memory profiler software	1	$5,000	http://www.b-alert.com/eeg/b-alert.html
Data Syncher	1	$2,500	http://www.seeingmachines.com/pdfs/brochures/faceLAB-5.pdf
FaceLab eyetracker	1	$32,500	http://www.seeingmachines.com/pdfs/brochures/faceLAB-5.pdf
Eyeworks Professional (software for FaceLab)	1	$5,900	http://www.seeingmachines.com/pdfs/brochures/faceLAB-5.pdf
FaceLab car/cockpit system	1	$2,500	http://www.seeingmachines.com/pdfs/brochures/faceLAB-5.pdf
Morae usability software	3	$4,500	http://www.techsmith.com/morae/getmorae.asp
Hitachi Medical fNIRS system	2	$370,000	http://www.hitachi-medical-systems.co.uk/index.php?id=29
interface test stations	3	$6,000	

The equipment purchased through this DURIP opportunity has benefited, and continues to benefit, research efforts between the Computer Science department at Hamilton College and colleagues working at the Air Force Office of Scientific Research, Air Force at Rome Lab, and at Wright Patterson Air Force Base. In the following section we provide an overview of the novel scientific research being conducted at Hamilton College, and we highlight the role that our

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **12 OCT 2011**	2. REPORT TYPE	3. DATES COVERED **01-06-2010 to 31-08-2011**

4. TITLE AND SUBTITLE **Hamilton College Usability Lab**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Hamilton College,198 College Hill Rd.,Clinton,NY,13323**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
With the help of DURIP funds, we created a new Usability Research lab at Hamilton College that enables us to gather a variety of cognitive and behavioral metrics from users in order to gather real time, quantitative measures of their mental states while they work with a variety of computer interfaces. The equipment purchased with the DURIP funds is listed in Table 1. Paramount to Hamilton?s approach is the use of a relatively new, non-invasive brain imaging technique called Functional Near-Infrared Spectroscopy (fNIRS) to make real-time, objective measurements of users? mental states while they work with various user interfaces. fNIRS was introduced in the early 1990?s to complement, and in some cases overcome practical and functional limitations of EEG and other brain monitoring techniques. Users can wear the comfortable fNIRS device in real working conditions, allowing the measurement of their brain activity while they work with a given interface. These unbiased, real-time metrics are used to augment more traditional usability tests, providing greater insight into the effects of a given interface design on target users.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**6**	

DURIP equipment has had in furthering research at Hamilton, and within the Air Force directorates with which we collaborate.

The novelty of our research stems from our use of a new, non-invasive brain measurement device to overcome the user state evaluation challenges listed above. We use our non-invasive brain measurements during our usability tests, where we have human users work with a target user interface (this interface could be to an airplane cockpit, an AFRL-sponsored web page, or any software developed for or by the Air Force, to name a few). While users work with the interface on prescribed tasks, we measure their brain activity, as well as a number of other, more traditional usability metrics. We then use the data from our usability experiment to evaluate the chosen interface design, to clearly specify its strengths (how it supports usability) and its weaknesses (how it limits usability), and to make suggestions about ways to improve (or further impair) the design.

We use EEG and fNIRS to acquire measures of users' mental states. Unlike other brain devices which require subjects to lie in restricted positions (fMRI), or to drink hazardous materials (PET), EEG and fNIRS can measure users' brain activity in realistic working conditions [1]. This makes EEG and fNIRS appropriate choices for brain measurement in usability testing. Whereas EEG has been available since the early 1900's, fNIRS was not introduced until the 1990's and holds great potential for extremely non-invasive cognitive state measurement. It is significantly easier and faster to set-up on subjects than is EEG, it is less sensitive to noise in the signal, has higher spatial resolution, and is less invasive, allowing for use in real working conditions. Our strong history of research with the promising fNIRS device has placed us at the leading edge of research on non-invasive measurement for usability testing and adaptive system design [2-6].

2 Review of Air Force Research

One of the strongest accomplishments that we reached since receiving our DURIP equipment was the creation, implementation, and validation of a novel usability experiment protocol and a set of machine learning methods that enable us to predict, on the fly, the user state of a given individual. Before we began this research, the **vast** majority of brain research (with the exception of some of the work by Berka et al [7]), and **all** fNIRS research could not PREDICT user states. This research could only predict that two (or more) user states differed from one another. We have had great success publishing our work in part because it offers a large leap forward in the state-of-the-art of non-invasive brain measurement in HCI. We are still working to sync up all of the devices in our lab and to build machine learning techniques that accepts concurrent recordings of data from all (or some subset of) the devices in our lab.

During the past year we also began research to quantify the user states of 'trust', 'distrust', and 'suspicion', that occur when a user works with a computer system. On a high level, our goal was to measure the cognitive (i.e. the fNIRS and EEG brain data), physiological (i.e., galvanic skin reponse, heartbeat, respiration), and behavioral metrics (i.e., keypresses, mouse movement, speed, accuracy) that occur when a user loses trust and becomes suspicious that his computer system has been hacked. The two primary AFOSR related end goals that stem from this basic research are:

1) ***Monitoring of Adversarial Computer Operators:*** It is worth noting that many of the behavioral metrics that we will measure can be measured remotely; meaning that we can simply record information on a user's computer, as opposed to requiring the user to wear some sort of measurement equipment. Currently, these remote metrics can be acquired in our usability experiments, and we can make controlled connections between the remote metrics and our concurrent recordings of cognitive and physiological data. In the future we could acquire these remote metrics directly from the computer of an adversarial computer operator. Therefore, by making ties to the remote metrics that occur while users are experiencing changes in their level of trust toward the computer, we can monitor adversarial computer operators and determine their level of trust toward their computer system—which we have hacked into (i.e., are they suspicious of our hack?). We can incorporate trust repair tactics as needed if we determine that the adversary has detected our security breach. This research has direct implications for the AFOSR Denial & Disrupt project, where we can run usability studies to determine the threshold at which certain interface disruptions become too overt, causing users to become suspicious of a computer hack.

2) ***Training Military Personnel***: By understanding the cognitive and emotional aspects of the trust relationship between a trainee and his computer system, we can better train our military personnel to eliminate cognitive shortcuts that sacrifice computer security in return for lower cognitive load and lower stress. Research suggests that people tend to be 'cognitive misers'. We often make cognitive shortcuts in order to minimize mental workload. One of these cognitive shortcuts includes adapting a *truth bias*, where people tend to believe that all people they interact with are being honest and truthful. The *truth bias* is a cognitive shortcut that keeps people from becoming suspicious of other people, as suspicion causes people to critically think about their interactions, causing a high level of cognitive load. We hypothesize that many personnel in the US military adapt this *truth bias* while working with their computer systems; it's simply too much effort for them to think critically about the validity of the information being presented in their interactions with their computer. Instead, they assume that they are usually not being hacked into, and they may miss a vital security breach. In the future, non-invasive cognitive, behavioral, and physiological measurement devices could be used in training sessions to ensure that military personnel develop the correct cognitive and physiological mental state to accurately detect security breaches.

With our strong research background and our cutting edge DURIP equipment, we have together with a team of experts in the trust and management domains. Our team of trust collaborators included Lt. Col Alex Barelka from AFRL at Wright Patterson AFB, Dr. Roger Mayer from NC State University, and Dr. Philip Bobko from Gettysburg College. We worked (and continue to work) to create new models of trust, distrust, and suspicion that describe the interactions between a person and that person's computer system. We also created plans to empirically validate our models and to measure a user's changing level of trust while working with his or her computer system.

Our goal within this larger team effort has been to find the cognitive, physiological, and behavioral correlates that predict users' level of trust, distrust, and suspicion toward their computer system. In addition to the Air Force work described above, we have also given tours of our lab to several members of the Air Force from Wright Patterson and to Air Force employees located at Rome Lab, the neighboring Air Force Base in upsate NY.

3 Other Air Force and DoD Research Efforts

With new funding efforts from the Army, to explore situational awareness in the brain, and from Booz Allen Hamilton, to gage the potential of using brain measurement as a biometric identification system, we feel well poised to continue using our DURIP lab to engage in interesting Air Force, and DoD related research in the future. We have also been pleased to coordinate our research efforts with those going on with Dr. Scott Galster and with Dr. Mike Haas' research groups at Wright Patterson AFB.

4 Publications Using DURIP Equipment

Finally, we have published, or are in the process of publishing, several research experiments that used the DURIP equipment in our lab [8-10].

5 References

1. Izzetoglu, M., et al., *Functional Near-Infrared Neuroimaging.* IEEE Trans Neural Syst Rehabil Eng, 2005. **13**(2): p. 153-9.
2. Hirshfield, L.M., *Enhancing Usability Testing with Functional Near Infrared Spectroscopy*, in *Computer Science*. 2009, Tufts University: Medford, MA.
3. Hirshfield, L.M., et al. *Combining Electroencephalograph and Near Infrared Spectroscopy to Explore Users' Mental Workload States*. in *HCI International*. 2009: Springer.
4. Hirshfield, L.M., et al. *Brain Measurement for Usability Testing and Adaptive Interfaces: An Example of Uncovering Syntactic Workload in the Brain Using Functional Near Infrared Spectroscopy*. in *Conference on Human Factors in Computing Systems: Proceeding of the twenty-seventh annual SIGCHI conference on Human factors in computing systems*. 2009.

5. Hirshfield, L.M., et al. *Human-Computer Interaction and Brain Measurement Using Functional Near-Infrared Spectroscopy*. in *Symposium on User Interface Software and Technology: Poster Paper*. 2007: ACM Press.
6. Girouard, A., et al., *From Brain Signals to Adaptive Interfaces: using fNIRS in HCI*, in *(B+H)CI: The Brain in Human-Computer Interaction and the Human in Brain-Computer Interfaces*, D.S. Tan and A. Nijholt, Editors. 2010, Springer.
7. Berka, C. and D. Levendowski, *EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning and Memory Tasks.* Aviation Space and Environmental Medicine, 2007. **78**(5): p. B231-B244.
8. Hirshfield, L., et al. *Trust in Human-Computer Interactions as Reflected by Workload, Frustration, and Surprise*. in *HCI International 2011 14th International Conference on Human-Computer Interaction*. 2011 (in press): Springer.
9. Hincks, S., et al. *The Engaged Brain: fNIRS-based Machine Learning Algorithms that Detect Flow in Video Games*. in *SIGCHI*. 2012 (submitted). Austin, Texas: ACM.
10. Escalante, J., L.M. Hirshfield, and S. Butcher, *Evaluating Interfaces Using Electroencephalography and the Error Potential*. 2012, Hamilton College.